Calculation of long time classical trajectories: Algorithmic treatment and applications for molecular systems

Hartmut Schwetlick[a] and Johannes Zimmer[b]
Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom

(Received 11 November 2008; accepted 18 February 2009; published online 23 March 2009)

We study the problem of finding a path that joins a given initial state with a final one, where the evolution is governed by classical (Hamiltonian) dynamics. A new algorithm for the computation of long time transition trajectories connecting two configurations is presented. In particular, a strategy for finding transition paths between two stable basins is established. The starting point is the formulation of the equation of motion of classical mechanics in the framework of Jacobi's principle; a shortening procedure inspired by Birkhoff's method is then applied to find geodesic solutions. Numerical examples are given for Müller's potential and the collinear reaction $H_2+H \rightarrow H+H_2$.
 [DOI: 10.1063/1.3096294]

I. INTRODUCTION

Complex systems in physics, chemistry, or biology can often be described by a potential energy landscape with many wells, separated by barriers. A common problem is then to find a trajectory joining a given initial point (configuration) with a given final point. We study this problem in the situation where the dynamics is determined by classical mechanics, and the given points are potentially far apart. In particular, the two configurations will generically be located in different wells of the energy landscape.

This is a classical problem, and a number of solution strategies have been proposed. We review some methods in Sec. II. For any numerical method, the question arises whether a computed numerical trajectory is close to a physical one, and if so whether it is close to a generic trajectory. Approximate orbits are said to *shadow* real trajectories if the answer to the former question is affirmative. Shadowing is known to hold for certain hyperbolic systems (e.g., Ref. 1). For molecular dynamics (MD), there is no proof of shadowing and thus computations are always based on trust.[2] Some evidence is given by Gillilan and Wilson,[3] where ideas related to the approach presented here are employed.

The focus on the boundary-value problem, where the initial and final states are given, is motivated by applications. Indeed, the initial and final points can be interpreted as reactant and product states. For problems of this kind, it is only theoretically possible to study the corresponding initial value problem, where the trajectory is completely determined by initial conditions for position and momenta. Namely, the given final point will only be reached by chance, even approximately. As described lucidly by Gillilan and Wilson,[3] the right choice of the initial momentum to reach a final destination is a difficult problem. Golfer can testify that this problem is challenging even in the relatively simple landscape of a golf course. For long-term trajectories in complicated potential energy landscapes, this difficulty is even more pronounced.

Rare events are an example where the efficient sampling of the configuration space is particularly important. Typically, thermally activated reactions have many deep wells separated by large energy barriers. Reactants will then spend most of the time jostling around in one well before a rare spontaneous fluctuation occurs that lifts the atoms of the reactant over the barrier into the next (product) valley. It is very difficult to choose the initial momentum so that this behavior can be observed. Yet, information on rare events is crucial since they represent important changes in the system, such as chemical reactions or conformational modifications of molecules. A major challenge in MD is that these hopping events take place so rarely that the computational limits of MD simulations can be easily exceeded.

Knowledge about transition paths, which connect two minima of the energy landscape, is important for path sampling techniques of statistical mechanics, where the aim is to determine transition states. Chandler and collaborators have notably advanced this technique;[4] the initial path is an important ingredient and thus the algorithm presented here has potential applications for transition path sampling as well.

Procedures to determine a realistic dynamical path for given initial and final positions often rely on equivalent variational formulations. Hamilton's principle of stationary action,[5] Maupertuis' principle,[6] and the Onsager–Machlup functional[7] are prominent examples. We rely on the Maupertuis principle as variational approach, namely, the representation of trajectories as geodesics in the Jacobi metric. We suggest a curve-shortening procedure which has some resemblance to rubber-band algorithms.[3] The method proposed here combines two crucial elements: (i) It is important that as variational formulation, Maupertuis' principle is chosen, as opposed to other, formally equivalent variational formulations. This is since we rely on a geometric argument that leads to curve shortening; this procedure, in principle, *converges* to a geodesic which represents a trajectory. A detailed

[a]Electronic mail: schwetlick@maths.bath.ac.uk.
[b]Electronic mail: zimmer@maths.bath.ac.uk.

mathematical analysis will be presented in a forthcoming publication. It has been stressed by developers of other approaches that those methods depend on choices such as approximating function spaces. The method presented here gives, in principle, arbitrarily small errors and thus may be of interest as a comparison method. (ii) Behind the geometric argument lies the observation that the underlying variational structure often leads to stationary solutions which are saddles rather than minima. This is a significant computational challenge, since saddle points are much harder to locate than minima. In Maupertuis' framework, a stationary curve (trajectory) can be represented as a concatenation of a number of minimal curves. We show that this property can be set to good use for numerical approximations. In principle, a curve-shortening (rubber-band) algorithm selects minima as solutions, while genuinely multiple solutions coexist, many of them being saddle points. We sketch how these points can be detected as well.

One benefit of the approach discussed here seems to be that reaction mechanisms can be found in an unbiased way; this should be of interest in application areas of MD, where the initial and final positions can, for example, be different configurations of a molecule. Another benefit is that reaction coordinates can be extracted from the paths obtained here.

This paper is organized as follows: In Sec. II, we introduce variational principles for the computation of trajectories and give a brief survey of some related work. Section II A is concerned with Hamilton's principle, while Sec. II B focuses on Maupertuis' principle. Section II B describes the setting and the theoretical framework for the algorithm presented here, and the method that leads to a rubber-band-like algorithm is summarized in Sec. III. Section IV describes the implementation in greater detail. Two applications are presented in Sec. V, and we close with a discussion in Sec. VI.

II. EQUATIONS OF MOTION AND RELATED STATIONARY PRINCIPLES

We review the variational principles of classical mechanics. Throughout the presentation, Q is the *configuration manifold* of a system and thus describes all possible states the system can occupy. The coordinates of the *phase space* T^*Q are (q^j,p_j), position, and momentum. Analogously, the coordinates of the tangent bundle TQ are $(q^j,\dot{q}^j)$, where $\dot{q}^j$ denotes the velocity. We assume that the system dynamics is conservative with $3N$ degrees of freedom. Then, the *Hamiltonian* $H:TQ\to\mathbb{R}$ is defined as $H:=E:=T+V$. Here, the kinetic energy $T=T(p)$ is a function of the momenta only and $V=V(q)$ is the potential energy, depending on the coordinates q alone. The *Lagrangian* of the system is a function $L:TQ\to\mathbb{R}$, namely, $L(q,\dot{q})=T-V$.

Before discussing different variational principles in detail, we remark that for chemical reactions with given reactant and product states, one often wants to find not only one but *all* dynamic paths joining these two states. Throughout this article, we write $\Omega(a,b,q_a,q_b)$ for the set of smooth curves $\gamma:[a,b]\to Q$, such that $\gamma(a)=q_a$ and $\gamma(b)=q_b$.

A. Hamilton's principle

Hamilton's principle states that every classical trajectory connecting two given configurations q_a and q_b is a stationary point for the action functional. The *action* is a functional $S:\Omega(a,b,q_a,q_b)\to\mathbb{R}$, given by

$$S[\gamma]:=\int_a^b L(\gamma(\tau),\dot{\gamma}(\tau))\mathrm{d}\tau. \tag{1}$$

Here, stationarity is understood in the following sense. We consider trajectories of our system, that is, curves in Q. More precisely, we consider curves $\gamma\in\Omega(a,b,q_a,q_b)$ for given q_a, $q_b\in Q$. Stationarity means being a critical point under variations,

$$\delta\int_a^b L(\gamma(\tau),\dot{\gamma}(\tau))\mathrm{d}\tau=0. \tag{2}$$

That is, for given initial and final states q_a and q_b, all dynamical paths correspond to stationary values of S. The classical equations of motions are the Euler–Lagrange equations for Eq. (1). Key facts of functional (2) are reviewed in Ref. 5; we only mention that this principle is a minimum principle only for nearby points, and never a maximum principle. Thus, for time scale of interest in MD, a challenge for a method relying on Hamilton's principle is that it is a saddle point principle. Algorithms for computing saddle points are computationally expensive and often only have a small radius of convergence. An approach by Passerone *et al.* which consists of augmenting Hamilton's principle via the addition of constraints to transform it into a minimum principle, is briefly reviewed in Sec. II C.

B. Maupertuis' principle

Historically older than Hamilton's principle is a variational principle that can be traced back to Maupertuis, which we call *Jacobi's least action principle*. This principle can be seen as a special case of Routh reduction applied to time as a cyclic variable (see, e.g., Ref. 8). We sketch the derivation for the sake of completeness. Let us consider an extended configuration manifold $\bar{Q}:=Q\times\mathbb{R}$ with coordinates (q^j,t), where t is time. Let $\langle\dot{q},\dot{q}\rangle=\Sigma_{j=1}^{3N}m_j\dot{q}_j\dot{q}_j$ be the inner product for a system with N particles with mass m_j. Then we assume that

$$L(q,\dot{q}):=\tfrac{1}{2}\langle\dot{q},\dot{q}\rangle-V(q) \tag{3}$$

is the original Lagrangian on TQ. We then define $\bar{L}:T(Q\times\mathbb{R})\to\mathbb{R}$ as $\bar{L}(q,q',t'):=L(q,q'/t')t'$, where the primed variables are differentiated with respect to a parameter τ, which is chosen to describe a curve on $\bar{Q}$. Then, with $\tau=\tau(t)$,

$$S[\gamma(\tau),t(\tau)]:=\int_a^b \bar{L}(\gamma(\tau),\gamma'(\tau),t'(\tau))d\tau, \tag{4}$$

$$= \int_a^b L\left(\gamma(\tau), \frac{d}{d\tau}\gamma(\tau)\frac{d\tau}{dt}\right)\frac{dt}{d\tau}d\tau = \int_{t(a)}^{t(b)} L\left(\gamma(t), \frac{d}{dt}\gamma(t)\right)dt. \tag{5}$$

Then the corresponding Euler–Lagrange equations are given as follows (D_j denoted differentiation with respect to argument j):

$$0 = \frac{d}{d\tau}D_2\bar{L}(\gamma(\tau), \gamma'(\tau), t'(\tau)) - D_1\bar{L}(\gamma(\tau), \gamma'(\tau), t'(\tau)), \tag{6}$$

$$\begin{aligned} 0 &= \frac{d}{d\tau}D_3\bar{L}(\gamma(\tau), \gamma'(\tau), t'(\tau)) \\ &= \frac{d}{d\tau}\frac{d}{dt'}\left[L\left(\gamma(\tau), \frac{\gamma'(\tau)}{t'}\right)t'\right] \\ &= \frac{d}{d\tau}\left[-D_2L\left(\gamma(\tau), \frac{\gamma'(\tau)}{t'}\right)\frac{\gamma'}{t'} + L\left(\gamma(\tau), \frac{\gamma'(\tau)}{t'(\tau)}\right)\right]. \end{aligned} \tag{7}$$

We note that the Euler–Lagrange Eq. (7) associated with the time variable is the energy conservation,

$$E = D_3\bar{L}(q, q', t') = D_2L\left(q(\tau), \frac{q'(\tau)}{t'(\tau)}\right)\frac{q'}{t'} - L\left(q(\tau), \frac{q'(\tau)}{t'(\tau)}\right) \tag{8}$$

for all τ. For Lagrangian (3), this constraint (8) can be expressed in the form

$$\frac{1}{2}\langle q', q'\rangle\frac{1}{(t')^2} + V(q) = E. \tag{9}$$

According to the Routh reduction procedure,[8] these Euler–Lagrange Eqs. (6) and (7) are also the Euler–Lagrange equations of the action constructed with the Routhian

$$R(q, q') := \bar{L}(q, q', t'(q, q')) + t'(q, q')E, \tag{10}$$

where the function $t'(q, q')$ is implicitly defined by constraint (9), namely,

$$t' = \sqrt{\frac{\langle q', q'\rangle}{2(E - V(q))}}. \tag{11}$$

We employ Eq. (11) to simplify Routhian (10) and obtain

$$R(q, q') = 2(E - V(q))\langle q', q'\rangle. \tag{12}$$

We remark that $R(\gamma, \gamma')$ is a metric in those regions of Q where $V(q) < E$. The action functional

$$J[\gamma] := \int_a^b R(\gamma, \gamma')d\tau \tag{13}$$

is the measure of the length of γ in this metric. For a given curve γ, the value $J[\gamma]$ is often called the *energy* of γ; the *length* of the curve is then

$$L[\gamma] := \int_a^b \sqrt{R(\gamma, \gamma')}d\tau. \tag{14}$$

The difference between length and energy is not very significant: if the energy functional is minimal, then the length functional is stationary as well. However, since the length functional is invariant under reparametrizations, reparametrizations of a curve of minimal length will be of minimal length as well, while a minimizer of the energy functional is automatically parametrized by arc length.

The previous considerations suggest that trajectories can be found as geodesics in Jacobi's metric (12). The *Maupertuis'* principle (sometimes denoted as *Jacobi's method*, or incorrectly, least action principle) relies on this observation; the derivation shows that it then results in the same equations of motion as Hamilton's principle. The total energy E is preassigned. Maupertuis' principle then seeks stationary solutions of functional (13) with metric (12). This principle is again a stationary one, which is why the popular name *principle of minimal action* is unfortunate.

We remark that the physical time can be recovered via the explicit formula

$$t = \int_0^\tau \sqrt{\frac{\langle q', q'\rangle}{2(E - V)}}ds. \tag{15}$$

Maupertuis' principle has been employed in a number of computational approaches and is regarded as a very accurate method suitable for the verification of other algorithmic formulations.[7] In particular, related to the method proposed here is the one developed by Banerjee and Adams.[6] While both methods are based on Maupertuis' principle, a major difference is that the approximation of Banerjee and Adams relies on a linear combination of basis functions. The choice of the approximating functions in Ref. 6 necessitates a global minimization, while we propose replacing a global minimization with a number of local minimizations. The method of Banerjee and Adams was acknowledged as useful for comparisons with other methods.[3]

C. Other approaches

Olender and Elber[7] based an algorithmic formulation on the Onsager–Machlup functional. Here a stationary point is found as a critical point of the norm of the derivative of the action functional. In practice, one first approximates the continuous action functional (1) by a discrete functional and then computes the norm of the derivative of the discrete (finite-dimensional) approximative action. For example, one can approximate a curve in Q by a finite set of points $q_a = q(0), \ldots, q(L) = q_b$. A reasonable approximation of action (1) is then

$$\sum_{l=1}^{L-1}\left[\frac{1}{2}m\left(\frac{q(l+1) - q(l-1)}{\Delta t}\right)^2 - V(q(l))\right]\Delta t. \tag{16}$$

Then the sum of the two norms (squares) of the derivatives of the discrete action functional is

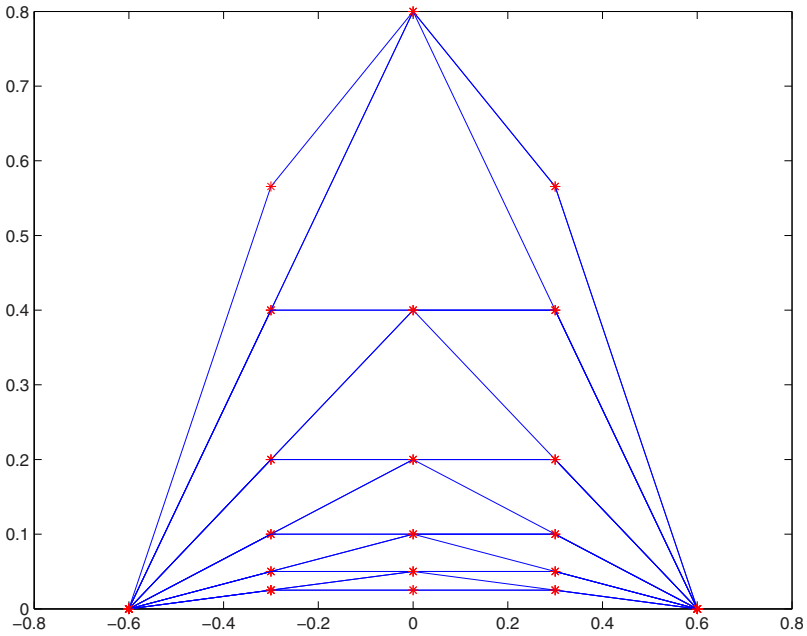

FIG. 1. (Color online) Birkhoff's algorithm, for the toy example of the Euclidean metric in $\mathbb{R}^2$. To find a geodesic joining the points $(-0.6,0)$ and $(0.6,0)$, we start with the outermost curve as initial guess. It is characterized by five points $q_0=(-0.6,0),\ldots, q_{2n}=(0.6,0)$ $(n=2)$. In a first step, the points with even indices are kept fixed, and joined by a geodesic. New positions for the points with odd indices on the new curve are determined. This results in the triangle-shaped curve with vertices $(-0.6,0)$, $(0,0.8)$, $(0.6,0)$. In a next step, the points with odd indices are joined by geodesics, which determine new positions for the points with even indices. The curves (slowly) converge to a geodesic line connecting q_0 and q_{2n}.

$$S_{\mathrm{OM}} = \sum_{l=1}^{L-1} \left[\frac{m}{\Delta t^2}(q(l+1) - 2q(l) + q(l-1)) - V'(q(l)) \right]^2 \Delta t. \tag{17}$$

This action coincides with a discretization of the Onsager–Machlup functional. A difficulty is that a numerical evaluation requires the computation of second derivatives, which is often not possible due to computational expenses.

Gillilan and Wilson[3] observed that the formulation of the Verlet algorithm for fixed end points can be interpreted in a variational manner. Furthermore, the methodology of Ref. 3 has an attractive interpretation, going back to the work of Peyrard and Aubry.[9] There, a connection between mappings and a model of a chain polymer adsorbed to a surface is established (see Sec. III). Gillilan and Wilson approximated a curve by discrete points $q(0),\ldots,q(L)$ with $q_a=q(0),\ldots,q(L)=q_b$. The discrete action reads

$$\sum_{l=1}^{L-1} \left[\frac{1}{2} m \left(\frac{q(l+1)-q(l)}{\Delta t} \right)^2 - V(q(l)) \right] \Delta t. \tag{18}$$

The discrete Euler–Lagrange equations for this functional are identical to the Verlet algorithm

$$q(l+1) = 2q(l) - q(l-1) - \frac{\Delta t^2}{m} V'(q(l)). \tag{19}$$

Passerone *et al.*[5] employed a hybrid Hamilton–Maupertuis iteration scheme. The key idea is that Hamilton's principle can, via the addition of constraints such as energy conservation as a Lagrange multiplier, be transformed into a minimum principle. The starting point is this modified Hamilton's principle. As input, a reasonable estimate of the reaction time $t:=b-a$ in Eq. (1) has to be provided. The resulting trajectory of the minimum principle is then used as input of Maupertuis' principle, which then results in a better estimate of the reaction time t so that the procedure can be iterated. This is a complicated and computationally demanding algorithm, but should provide a good basis for comparisons.

III. MAUPERTUIS' PRINCIPLE AND CURVE SHORTENING

The main difference between existing approaches and the method proposed here is that we rely on a well-known and simple observation: R of Eq. (12) defines a metric. Consequently, solutions of the equations of motions are geodesics in metric (12), where the geodesic variational formulation is given in Eq. (13). This immediately shows that Maupertuis' principle is plagued by the same difficulty as Hamilton's principle. Indeed, geodesics are in general not minimizers, even if compared to nearby curves. For example, a segment of a great circle on the unit sphere that includes the north and south poles as interior points of the curve is a geodesic. However, length functional (14) for such an arc is stationary and not minimal; the same is true for the energy functional (13).

Yet, there is one simple, but significant fact: while geodesics are themselves not necessarily length minimizing, every segment of a geodesic that is short enough minimizes length functional (14) [and simultaneously the energy functional (13)] for given initial and final points. Thus even geodesics which are not minimizing can be thought of as a sequence of minimizing geodesics glued together. Birkhoff took advantage of this property which forms the basis of the *Birkhoff curve-shortening algorithm*. We outline the idea (see Fig. 1). For a given initial and final position q_a and q_b, consider an arbitrary curve in the configuration manifold Q connecting these two states. To make R a metric, we require this curve to consist only of points q where $V(q)<E$. The curve is partitioned in $2n$ segments, where $n\in\mathbb{N}$ is chosen such that the segments are small enough to be length minimizing. The end points of the segments are $q_0,\ldots,q_{2n}$. Then in a first step, neighboring points with even indices are

joined by a geodesic. See Fig. 1. Since the geodesic is length minimizing, the total length of the new curve cannot exceed the length of the initial curve. For the new curve, one can define points q_{2k+1} and join these points in a second step by geodesics. It can be shown that if this procedure is repeated, the approximations obtained in this way converge to a geodesic.

Given that we interpret geodesics as dynamic trajectories, this theoretical result seems remarkable since it results in convergent approximations. This seems important in the light of the open problem as to whether numerical trajectories shadow physical trajectories. However, we point out that the convergence of the Birkhoff curve-shortening algorithm relies on the computation of local geodesics between next to nearest neighboring points. Numerical approximations will inevitably introduce numerical errors. It is then no longer guaranteed that a step in the algorithm will indeed shorten the length of the curve. We thus describe here a suitable numerical implementation and investigate the feasibility. A theoretical study of the convergence of curve-shortening algorithms will be the topic of a separate investigation.

It is noteworthy that Birkhoff's curve-shortening algorithm requires (in theory) no refinement; both input and output are $2n$ points approximating a curve, where $n \in \mathbb{N}$ is kept fixed. (This requires n to be large enough since otherwise the shortening can produce points which have shorter radius of injectivity than their preimages.) Also, the method is self-correcting in the sense that if the geodesic segments are not calculated correctly, then this corresponds to a modification of the initial curve; the iteration will take longer, but will eventually converge, unless further systematic errors are introduced in the subsequent iterations.

Birkhoff's curve-shortening algorithm is related to so-called rubber-band algorithms. We discuss this first for the analysis of the Verlet algorithm from a variational perspective by Gillilan and Wilson[3] (see Sec. II C). We recall that a curve is approximated there by discrete points $q(0)$ with $q_a = q(0), \ldots, q(L) = q_b$. This discretization of the kinetic energy is, following Peyrard and Aubry,[9] interpreted as an elastic restoring force between neighboring interpolation points. It is the force related to the elastic energy

$$\sum_{l=1}^{L-1} \left[\frac{1}{2} m \left(\frac{q(l+1) - q(l)}{\Delta t} \right)^2 \right] \Delta t \tag{20}$$

between neighboring interpolation points. The elastic force then reads in suitable units

$$[q(k+1) - q(k)] - [q(k) - q(k-1)],$$

it is a harmonic force that balances the force V, which in the setting of Peyrard and Aubry, originates from the surface. Specifically, the point (bead) $q(k)$ is pulled to the right by $q(k+1)$ while simultaneously being pulled to the left by $q(k-1)$. This elastic restoring force is counterbalanced by the force $V'(q)$; the negative sign in front of the potential energy V in the action functional results in pushing the interpolants *uphill* with respect to the potential function. Thus, the kinetic energy, interpreted as elastic potential, exactly balances the negative potential energy at every discretisation point. The instability of the saddle point nature is also reflected by the opposite signs of the elastic potential and the negative potential energy $-V(q(l))$ in Eq. (18). A change of the sign of V consequently yields a stable situation, where Eq. (18) has a minimum. This change of sign of $V(q)$ results in what is denoted the elastic band method.

A similar viewpoint is taken by Czerminski and Elber,[10] who presented various methods of regarding reaction pathways as a chain of configurations arranged over a hilly landscape including mountain passes, where the chains are then pulled taut. This approach is in spirit very similar to the algorithm presented here. We develop in this paper a systematic method to pull taut, and it turns out that Maupertuis' principle is a particularly suitable formulation.

IV. NUMERICAL IMPLEMENTATION

The algorithmic difficulty of Birkhoff's curve-shortening method is that here metric (12) is not Euclidean. Geodesics are thus not straight lines so the implementation becomes more complicated. However, there is a rather natural discretization which mimics the trivial Euclidean algorithm. Namely, almost every algorithm will represent a given curve γ by a set of points $q(l)$ on the curve, with $l = 1, \ldots, L$. If L is chosen large enough, then the piecewise interpolation of these points will be a reasonable approximation of the curve. This is the approximation we make in step 1 in the algorithm below. It is now natural to discretize the metric g as well; we introduce a piecewise constant approximation in step 2 below. Specifically, we choose the approximative metric in a suitable neighborhood of a point $q(l)$ such that it coincides with the metric $g(q(l))$ of point $q(l)$. Now the situation is piecewise Euclidean. The computation of the length of the curve is very simple, and implementations of curve-shortening algorithms such as Birkhoff's algorithm are straightforward.

Before discussing the algorithm in detail, we remark that there is a theoretical foundation of the approach taken here. Namely, it is not hard to show that piecewise Euclidean metrics are dense in the set of Riemannian metrics (in fact even in a larger class, namely, Finsler metrics).[11]

A sketch of the method is as follows:

(1) Input: an initial polygonal curve γ joining given configurations q_a and q_b, as sketched in Fig. 2. The curve has to lie inside the configuration manifold, $\gamma \subset Q$. We represent γ by its nodes $q(l)$, $l = 1, \ldots, L$, with $q_a = q(0), \ldots, q(L) = q_b$. Here, L has to be chosen to be sufficiently large.

(2) Approximate the metric R of Eq. (12) in a suitable neighborhood of the curve by a piecewise Euclidean metric, such that the metric in the vicinity of a discretisation point $q(l)$ is given by the metric of that point, $g(q(l)) = E - V(q(l))$. See Fig. 2. We now consider the straight segment joining $q(l-1)$ with $q(l)$. The normal line through the midpoint of this segment is the interface where the approximate Euclidean metric will jump from $g(q(l-1))$ to $g(q(l))$, see Fig. 2. The boundary between $q(l)$ and $q(l+1)$ is defined analogously. We point out that this piecewise metric is not defined glo-

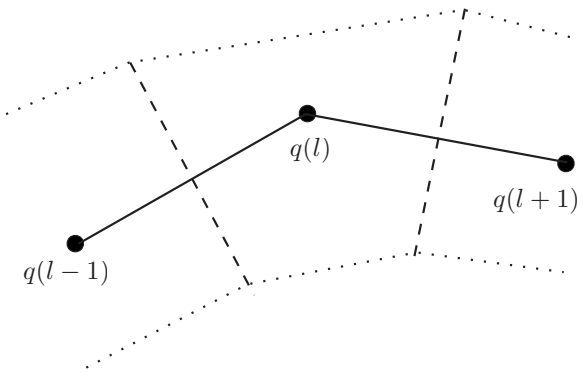

FIG. 2. Sketch of the algorithm's steps 1 and 2. Shown in the polygonal curve between $q(l-1)$ and $q(l+1)$ (solid line); the solid and the dotted lines define three regions. For the region surrounding $q(l)$, we choose $g(q(l))$ as metric for the entire region, and analogously for $q(l-1)$ and $q(l+1)$. Thus, in each region, the metric is Euclidean.

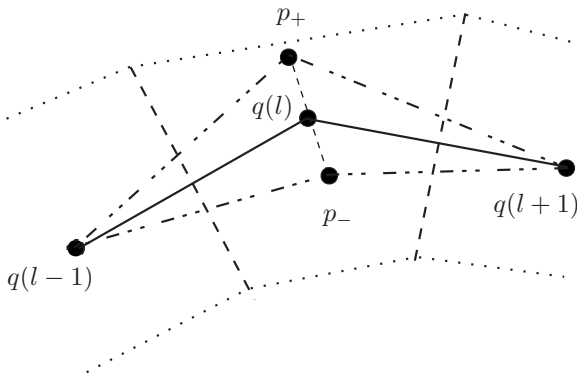

FIG. 3. Sketch of the algorithm's steps 3 and 4. In addition to the objects shown in Fig. 2, the points $p_{\pm}$ are indicated; the vector ν defines the line joining p_+ and p_-. The curves joining $q(l-1)$, $p_{\pm}$ and $q(l+1)$ are shown as dashed-dotted lines.

bally; however, for the algorithm it suffices to define the metric only in the vicinity of the given curve.

(3) Iteratively single out three neighboring discretisation points and move the middle one to decrease length. This captures Birkhoff's idea; in practice, a flow model is better suited and is described in detail below.

(4) Jacobi's metric is not constant and degenerates at the boundary. Thus under iteration, neighboring points can come close to each other faster than neighboring points further away from the boundary $V(q)=E$. This normally leads to an increasingly uneven distribution of points during the iteration. We thus reparametrize the curve by arc length to avoid a clustering of discretization points.

Some details are in order. In step 3, we proceed as follows. As in Birkhoff's algorithm, given three points, we want to move the middle one to shorten the length. We observe that it suffices to search for candidate points for the new middle point in all normal direction to the curve. We thus define a normal to the polygonal curve. Let us write $\tau_- := q(l)-q(l-1)$ and $\tau_+ := q(l+1)-q(l)$ for the discretized tangent vectors. The mean $\tau := \frac{1}{2}(\tau_- + \tau_+)$ is then an averaged discrete tangent vector at $q(l)$. Let ν be a normalized unit vector orthogonal to τ. We then consider the flow, that is, the motion of $q(l)$ in the direction of the normal ν. Suppose for the moment that all quantities are continuous, rather than discretized. Then, since the gradient $\nabla J(q(l))$ points in the direction of the strongest increase in J, the gradient flow in direction of ν directs us toward the strongest decrease of J in direction ν,

$$\frac{\partial}{\partial s} q(l) = -\nabla J(q(l))\nu. \tag{21}$$

We thus mimic this flow in a discrete setting. First, we approximate the gradient of the length functional as follows. For $\delta>0$, we consider the points $p_{\pm} := q(l) \pm \delta\nu$, see Fig. 3. Let J_+ be the energy of the segment consisting of two lines joining $q(l-1)$ and $q(l+1)$ via p_+. Here, J_+ is the energy with respect to the piecewise Euclidean metric of Fig. 2. Analogously, J_- is the energy of the segment connecting $q(l-1)$ and $q(l+1)$ via p_-. Then $(J_+-J_-)\nu$ is a discrete approximation of $\nabla J(q(l))\nu$. Second, we approximate the time derivative $(\partial/\partial s)q(l)$ by a difference quotient $(1/\Delta\xi)[q(l)^{\text{new}} - q(l)]$ for a constant $\Delta\xi>0$. Then the new position of $q(l)$ is determined by the discrete version of the gradient flow (21),

$$q(l)^{\text{new}} := q(l) - \Delta\xi \cdot (J_+ - J_-)\nu. \tag{22}$$

We recall that this discretization step involves two parameters, δ and $\Delta\xi$. The latter introduces an artificial evolution time s, whereas the former is related to the aspect ratio of the cell $(q(l-1), p_-, q(l+1), p_+)$ (dashed-dotted in Fig. 3). The aspect ratio of this cell should not degenerate under iteration. We thus have to choose $\delta = O(\|\tau\|)$, say $\delta = d\|t\tau\|$ with $d=\frac{1}{4}$.

Analogously to the explicit discretisation of parabolic equations, we expect a criterion similar to the Courant–Friedrichs–Lewy condition imposing an upper bound to the choice of $\Delta\xi$. Indeed, since the difference J_+-J_- of the path lengths scales like $2\delta\cdot\|t\tau\| = 2d\|t\tau\|^2$, we see that

$$\Delta\xi = \frac{\Delta s}{2d\|\tau\|^2}$$

can be interpreted as a Courant–Friedrichs–Lewy number for the discretization of the continuous flow Eq. (21).

As for step 4, we interpolate the new set of points with a cubic spline. For this curve, the notion of arc length with respect to the Euclidean metric is well defined, and it is straightforward to redistribute the points along the spline curve so that they are spaced in an equidistant manner. Here, it is better to use the Euclidean metric, since the Jacobi metric or the piecewise Euclidean approximation degenerate at the boundary. This would lead to an uneven distribution of points, in particular, close to the boundary. Furthermore, the computation of the piecewise Euclidean metric depends by construction on the choice of the points $q(l)$; reshuffling these points would thus change the metric and require us to recompute it, making the task unnecessarily expensive.

The explanations above, especially Figs. 2 and 3, illustrate the algorithm in two space dimensions. The extension to high space dimensions is straightforward; the only change is that the descent in normal direction is now in a higher-dimensional space. Specifically, for any dimension $n \in \mathbb{N}$, the normal plane to a tangent through a point is $n-1$-dimensional hyperplane. Given that we search at a

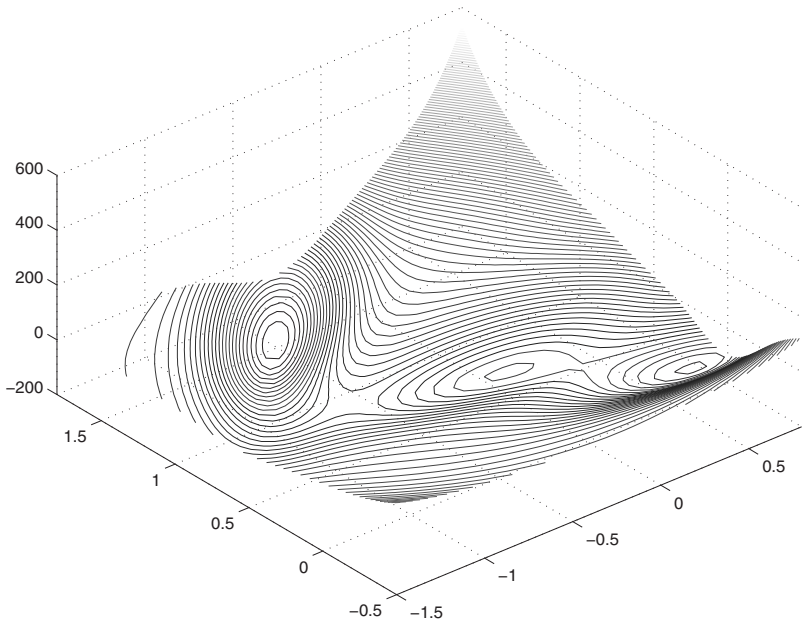

FIG. 4. Contour plot of the Müller potential.

given distance from $q(l)$, this leaves us with a $n-1$-dimensional sphere to search for the optimal direction. For example, in $n=2$, the sphere consists of two points, $J_{\pm}$. Thus, the algorithm has the promising feature that it scales linearly in the dimension of the problem.

We remark that one needs to impose an algorithmic constraint for the curve to avoid the region $E<V$ since for such points the weighted length $\sqrt{E-V}$ ceases to be defined. We ensure this by computing $E-V$ first; in the computation of the square root a negative value of $E-V$ is replaced by a large multiple of its absolute value. Thus, points which are outside the region $E-V$ will be rapidly dragged back by the length-shortening procedure described above. It is important to notice that the method presented here does not require the computation of second derivatives of the potential energy, as it is the case for methods based on the Onsager–Machlup principle.

A. Minimality versus stationarity

As pointed out by Jacobi and explained at the beginning of Sec. III, geodesics are in general not minima. In a rubber-band visualization, a geodesic on a sphere can be thought of as a rubber band stretched along an arc of a great circle of the sphere. If the arc covers more than half of the circumference, then it is a saddle point and a suitable small perturbation will make the rubber-band slide until it becomes the minimal configuration for the given initial and final point, that is, the "short" segment of the great circle joining the points.

Trajectories are commonly saddle points, which means they are minima under some perturbations and maxima with respect to other perturbations. It is difficult to find *all* dynamical trajectories connecting two given states. The method described here detects minimal curves (geodesics). However, it is possible to use the algorithm to approximate saddle points as well. This feature is understood intuitively with the picture of the rubber-band interpretation in mind. Suppose a rubber band connecting beads is in an uphill or downhill region, but not in a minimum or a saddle point. Then pulling the band taut will result in relatively quick changes of the position of the beads. However, when crossing a saddle point, genuinely some part of the rubber band will be pulled uphill and others downhill. As a consequence, the overall length and energy of the curve do not change abruptly due to compensation effects. Eventually, the band will be pulled over the saddle point, but it may take considerable time to do so. Thus, candidates for saddle points can be detected by investigating regions where the energy decreases slowly. In Sec. V, we give an example of a saddle point connection found on the basis of curve shortening.

V. APPLICATIONS AND NUMERICAL EXAMPLES

Thermodynamic and kinetic properties of molecular structures are linked to their potential energy surface (PES). PESs are normally nonconvex due to the presence of many minima. Minima are often separated by high energy barriers, and the configurational space can be disconnected. The energy landscape is often high-dimensional and highly complicated. We do not claim that we can presently tackle these challenging problems; at the moment, we aim to investigate the feasibility of the method by analyzing well-established model problems. Both are two-dimensional problems; the first one, the Müller potential, is discussed in Sec. V A, while the reaction $H_2+H \rightarrow H+H_2$ is the topic of Sec. V B.

A. The Müller potential

The Müller potential[12] is a common nontrivial test case for transition path methods (see, e.g., Refs. 7 and 13). The potential surface (see Fig. 4) has three minima, with the global minimum of -146.7 located at $(-0.558,1.442)$ $(-0.558,1.442)$. The Müller potential has two degrees of freedom and involves the parameters $A=(-200,-100,-170,15)$, $a=(-1,-1,-6.5,0.7)$, $b=(0,0,11,0.6)$, $c=(-10,-10,-6.5,0.7)$, $x_j=(1,0,-0.5,-1)$, and $y_j=(0,0.5,1.5,1)$; the analytic form of the Müller potential is then

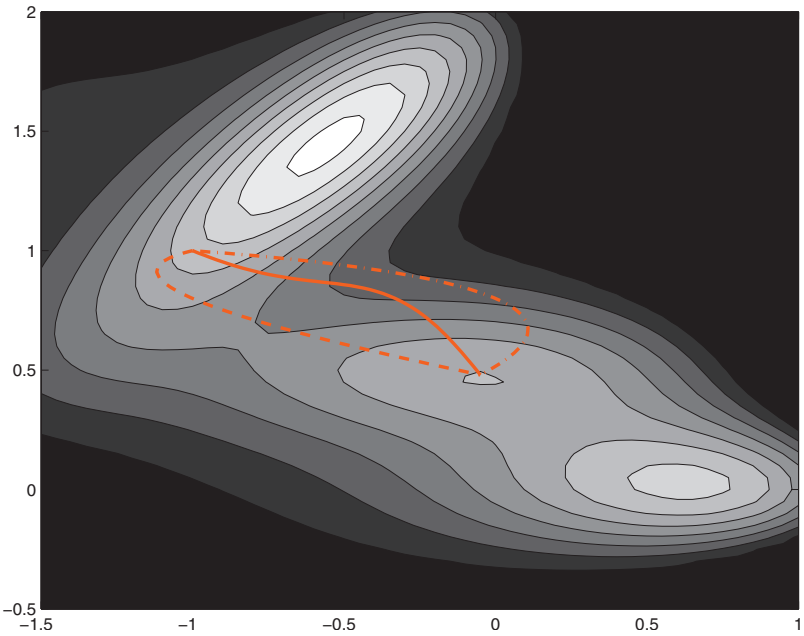

FIG. 5. (Color) Two initial curves (lower dashed and upper dashed-dotted curves) approaching the same minimum (solid trajectory) of the length functional for the Müller potential with $E=20$.

$$V(x,y) = \sum_{j=1}^{4} A_j \exp[a_j(x-x_j)^2 + b_j(x-x_j)(y-y_j) + c_j(y-y_j)^2]. \quad (23)$$

Despite the simplicity of this expression, the reaction coordinates are highly contorted, see Fig. 4 for a plot of the Müller potential. Figure 5 shows a trajectory joining two given points; here $E=20$, and we employ the algorithm to compute paths connecting the points $(-1,1)$ and $(0,0.5)$. The outer dashed and the outer dashed-dotted initial curve both converge to the solid curve in the middle, which is a local minimum of the length functional and thus a trajectory joining the two points.

We remark that the algorithm can be used to approximate unstable stationary curves as well. In combination with a bisection procedure it is possible to find saddle connections. This is shown in an example with the Müller potential, see Fig. 6. A detailed analysis of the detection of saddle points will be presented in a forthcoming publication.

B. The collinear reaction ($H_2+H \rightarrow H+H_2$)

As a second example, we consider the gas phase exchange reaction $H_2+H \rightarrow H+H_2$. This reaction, together with its isotropic variants, is the simplest of chemical reactions and has served for about 70 years as a common testing ground for computational studies of chemical reaction kinetics. The atom-diatom reaction of $H+H_2$ is completely captured by two coordinates R_1 $(=R_{12})$ and R_2 $(=R_{23})$. Banerjee and Adams[6] and Dey *et al.*[14] studied algorithms based on

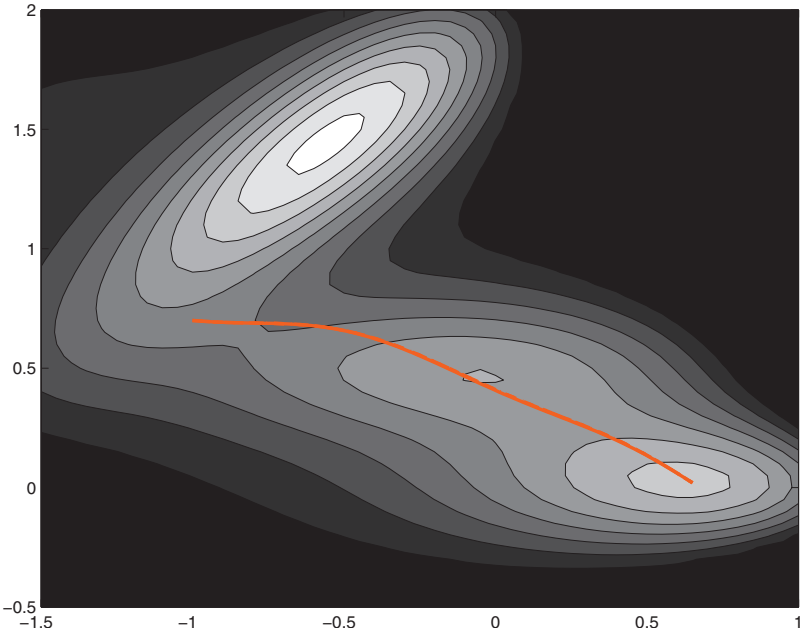

FIG. 6. (Color) A saddle connection of two points for the Müller potential.

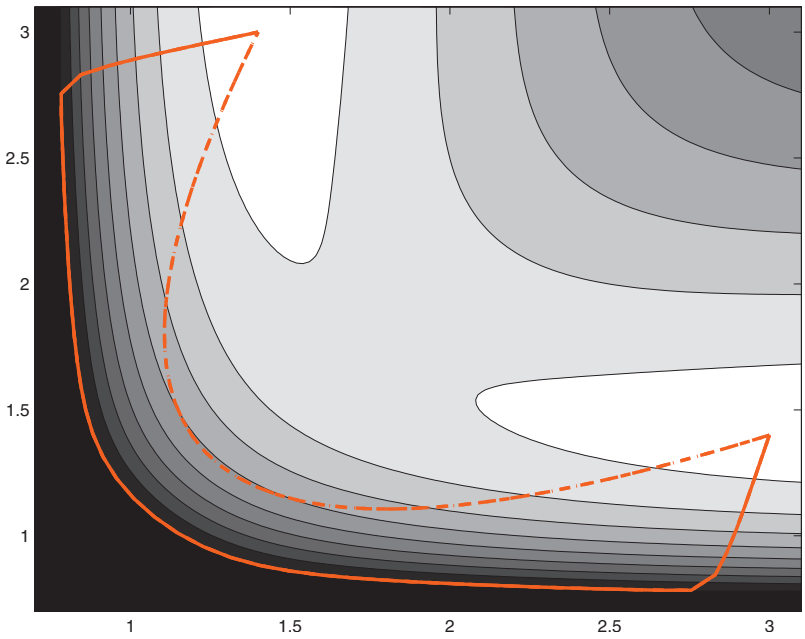

FIG. 7. (Color) The initial curve (dashed) and its corresponding local minimum of the length functional (solid) for the H_3 potential with $E=-1.5$ a.u.

Maupertuis' principle and the Hamilton–Jacobi approach, respectively, with this reaction. In both cases, the potential energy landscape for linear H_3, which has the asymptotic limit $H_2+H\rightarrow H+H_2$, is taken from Liu.[15] We use this potential as well. Let us write $[k/2]$ for the largest integer less or equal to $k/2$; the potential energy is then given as

$$V(R_1,R_2) = V_0 + \sum_{j=1}^{2} V_{H_2}(R_j) + \exp(-\gamma(R_1+R_2)) \times \sum_{k=0}^{n} \sum_{j=0}^{[k/2]} C_{k-j,j}(R_1^{k-j}R_2^j + R_1^j R_2^{k-j}), \tag{24}$$

with the potential energy V_{H_2} of H_2 given by

$$V_{H_2}(R_j) = -1 + \exp\left(-\alpha R_j \sum_{l=0}^{m} a_j R_j^l\right). \tag{25}$$

Here the parameters read $n=14$, $m=8$, $V_0=0.5$, $\alpha=1.914\,062\,5$, and $\gamma=1.539\,062\,5$; whereas C_{jk} and α_j are taken from Ref. 15, Tables VIII and III respectively. The potential has a saddle point where the energy is 9.8 kcal/mol above the (calculated) energy of an isolated hydrogen atom H and a hydrogen molecule H_2, and 10.28 kcal/mol above the exact energy of the system. We remark that the potential is symmetric, $V(R_1,R_2)=V(R_2,R_1)$. Furthermore, $\lim_{R_2\rightarrow\infty} V(R_1,R_2)=-0.5+V_{H_2}(R_1)$.

In the first computations we prepare the system with finite energy $E=-1.5$ a.u., which is slightly above the saddle point energy $E=-1.658$ a.u. The reactant and product con-

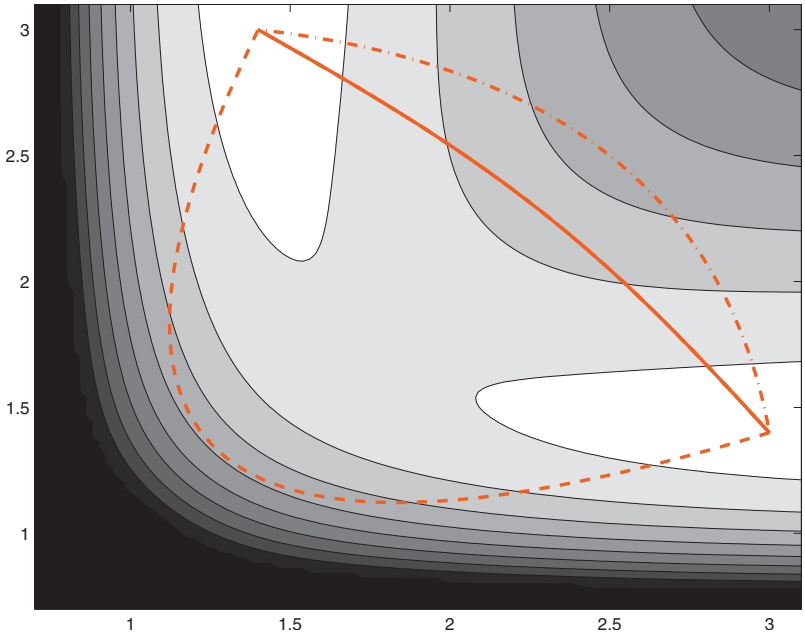

FIG. 8. (Color) Two initial curves (lower dashed and upper dashed-dotted curves) and their corresponding local minimum (solid curve) of the length functional for the H_3 potential with $E=-1.5$ a.u. The solid curve describes a trajectory joining the initial and final state.

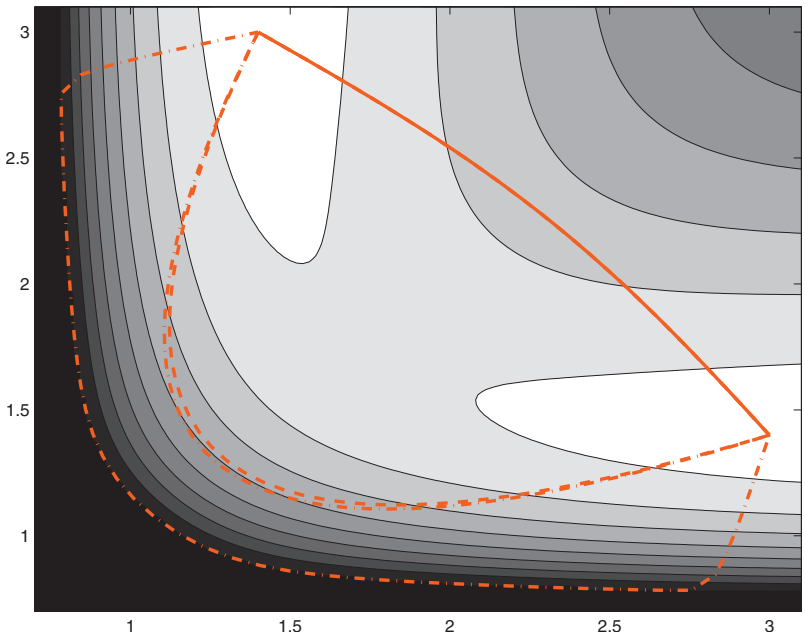

FIG. 9. (Color) The separation of two domains of attraction for the H_3 potential with $E=-1.5$ a.u. While the upper dashed curve converges to one local minimum (solid curve), the nearby dashed-dotted curve converges to different local minimum along the boundary (outer dashed-dotted curve). The thin region between the dashed and the dashed-dotted initial curves approximates the position of the saddle curve separating the two domains of attraction.

figuration are chosen to be near the entrance channel $R_1=3$ bohrs and $R_2=3$ bohrs for the reactant and the symmetric state $R_2=3$ bohrs and $R_2=3$ bohrs for the product state. Similar values have been chosen for an approximation of the critical trajectory by finite-dimensional approximation.[16]

In Figs. 7 and 8, we describe how the algorithm can be applied to find local minima of the length functional for H_3. Figure 7 shows a given initial curve dashed). This is then deformed toward a local minimum (solid curve). A simulation with identical parameters, but a different initial curve can lead to different minima, as shown in Fig. 8. Here the two different initial curves (the dashed and dashed-dotted curves) both converge to the inscribed curve (solid curve). In fact, in this special two-dimensional example the algorithm obeys a comparison principle, implying that all curves initially inscribed by the two outer curves will converge to the very same minimum.

To picture the different domains of attraction, we superimpose the two figures in Fig. 9. We changed the line style so that the dashed-dotted initial curve converges to the dashed-dotted local minimum whereas the dashed curve converges to the solid local minimum. We remark that the two initial curves lie very close together, which indicates that we can expect a saddle connection in this region.

To compute such saddle connections numerically, we employ the algorithm described above in conjunction with a bisection procedure. Here initial curves are deformed until they enter the regions already identified as domains of attraction of the local minima. We remark that while in principle this approach extends to higher dimensions, the two dimensionality of the test problem greatly facilitates the analysis. It is noteworthy that, in this way, we approximate curves that lie in the stable manifold of the saddle. An iteration of the procedure then approaches the saddle curve. The bisection

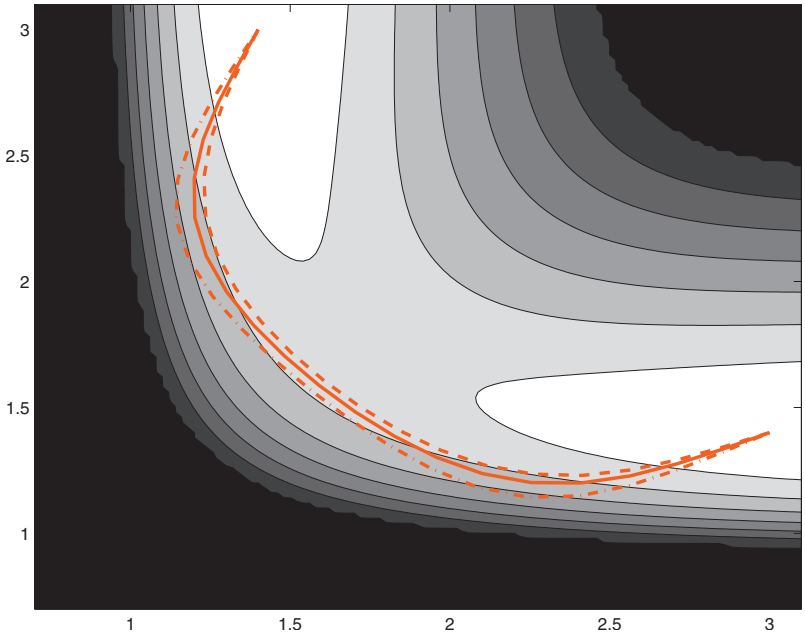

FIG. 10. (Color) A saddle connection for the H_3 potential with $E=-1.6$ a.u.

algorithm is shown in Fig. 10 for $E=-1.6$ a.u., which is slightly closer to the saddle point energy. The solid curve in the middle is an approximation for the saddle whereas the dashed and the dashed-dotted curve on either side belong to the two different domains of attraction and will converge eventually to their corresponding local minima.

VI. CONCLUSIONS AND PERSPECTIVES

As alternative rubber-band formulations, the method presented here transform the dynamical problem describing a trajectory into a static one. The algorithm presented here scales only linearly in the dimension of the problem, which seems promising in particular in comparison to other methods, where approximations in function spaces scale less favorably.

As most related methods (e.g., Ref. 6), the approach presented here requires an initial guess of a trajectory as an input, which has to be contained in the region where $E-V(q)\geq 0$. It seems an advantage that once this input (and the total energy E and the potential V, of course) is provided, no further information is required. This is different from the approach of Ref. 6. There, critical points of a finite-dimensional approximation of the action functional are determined. It is noted that the choice of the approximation, such as hyperbolic functions for dissociating molecules, is crucial. No such choice is required from the user with the method introduced in the present paper.

It is remarkable that it is not known rigorously whether shadowing holds for molecular trajectories or not. There is supporting evidence for the variational Verlet algorithm in some cases,[3] but there is no mathematical proof (the mathematical analysis is confined to relatively simple model problems such as the driven pendulum). The mathematical framework of shadowing relies on uniform hyperbolicity which can roughly be thought of as ergodicity. Thus, in essence, ergodicity has to be assumed if shadowing should hold. The approach presented here seems remarkable in the sense that no such assumption is required for the algorithm discussed above. At its core, it is an implementation of the Birkhoff curve shortening which converges under weak assumptions. A detailed mathematical analysis will be the subject of a forthcoming publication.

ACKNOWLEDGMENTS

J.Z. gratefully acknowledges the financial support of the EPSRC through an Advanced Research Fellowship (Grant No. GR/S99037/1). Both authors benefited from helpful discussions during the first annual meeting of the EPSRC network "Mathematical Challenges of Molecular Dynamics: A Chemo-Mathematical Forum" (Grant No. EP/F03685X/1).

[1] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia of Mathematics and its Applications, Vol. 54 (Cambridge University Press, Cambridge, 1995).
[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Computational Science Series, 2nd ed., Vol. 1 (Academic, Orlando, FL, USA, 2002).
[3] R. E. Gillilan and K. R. Wilson, J. Chem. Phys. **97**, 1757 (1992).
[4] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[5] D. Passerone, M. Ceccarelli, and M. Parrinello, J. Chem. Phys. **118**, 2025 (2003).
[6] A. Banerjee and N. P. Adams, J. Chem. Phys. **92**, 7330 (1990).
[7] R. Olender and R. Elber, J. Chem. Phys. **105**, 9299 (1996).
[8] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, Texts in Applied Mathematics, 2nd ed., Vol. 17 (Springer-Verlag, New York, 1999).
[9] M. Peyrard and S. Aubry, J. Phys. C **16**, 1593 (1983).
[10] R. Czerminski and R. Elber, Int. J. Quantum Chem. **38**, 167 (1990).
[11] A. Braides, G. Buttazzo, and I. Fragalà, Asymptotic Anal. **31**, 177 (2002).
[12] K. Müller, Angew. Chem., Int. Ed. **19**, 1 (1980).
[13] D. Passerone and M. Parrinello, Phys. Rev. Lett. **87**, 108302 (2001).
[14] B. K. Dey, M. R. Janicki, and P. W. Ayers, J. Chem. Phys. **121**, 6667 (2004).
[15] B. Liu, J. Chem. Phys. **58**, 1925 (1973).
[16] A. Banerjee, N. P. Adams, and C. Clawson, Chem. Phys. Lett. **168**, 149 (1990).